Journal of Water and Soil
https://jsw.um.ac.ir

Research Article
Vol. 38, No. 3, Jul.-Agu. 2024, p. 367-382

Evaluating the Precision of the Conditioned Latin Hypercube Sampling Method for Selecting Soil Samples to Generate Digital Maps of Soil Properties

Z. Mosleh Ghahfarokhi[1*], A. Azadi[2]

1- Assistant Professor of Soil and Water Research Department, Chaharmahal and Bakhtiari Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Shahrekord, Iran
(*- Corresponding Author Email: Z.mosleh@areeo.ac.ir)
2- Assistant Professor of Soil and Water Research Department, Khuzestan Agricultural and Natural Resources Research and Education Center, AREEO, Ahvaz, Iran

Received: 19-02-2024 Revised: 03-04-2024 Accepted: 07-05-2024 Available Online: 07-05-2024	**How to cite this article:** Mosleh Ghahfarokhi, Z., & Azadi, A. (2024). Evaluating the precision of the conditioned latin hypercube sampling method for selecting soil samples to generate digital maps of soil properties. *Journal of Water and Soil, 38*(3), 367-382. (In Persian with English abstract). https://doi.org/10.22067/jsw.2024.86936.1388

Introduction

Soil properties play a crucial role as they determine the soil's suitability for different types of plant growth, ecosystems, and biota functioning. They have a significant impact on nutrient cycling, carbon sequestration, and soil management. Digital Soil Mapping (DSM) is a process aimed at delineating soil properties. Soil sampling for DSM serves as a fundamental step in improving prediction accuracy and is crucial for incorporating variability in terms of environmental covariates. Conditioned Latin Hypercube (CLH) sampling is a technique utilized to generate a sample of points from a multivariate distribution conditioned on one or more covariates. Numerous researchers (Ramirez-Lopez *et al*., 2014; Adhikari *et al*., 2017; Zhang *et al*., 2022) have endorsed this approach in their studies, following its inception by Minasny and McBratney in 2006. However, there has been limited research to date on the impact of the Latin hypercube method's random sample selection process on the accuracy of resulting maps. Hence, the central question remains: Is the Latin hypercube sampling method, which is currently widely adopted, always a dependable approach in this field?

Materials and Methods

The study area covers longitudes 50°35'47" to 51°29" east and latitudes 31°36"31" to 32°15'48" north in Borujen city, Chaharmahal, and Bakhtiari Province. The region, with an average elevation of 2338 meters above sea level, receives an annual rainfall of 250 millimeters and maintains an average temperature of 11.5 degrees centigrade. In this investigation, inherited data from soil studies were utilized, consisting of 250 samples distributed across the study area. In this research, the studied characteristics included percentage of equivalent calcium carbonate, clay, and soil organic carbon at a depth of 0 to 30 cm. Land component variables were extracted using the Alus Palsar digital elevation model with a spatial resolution of 12.5 meters. In the initial stage, digital maps of equivalent calcium carbonate, clay, and soil organic carbon were generated using the support vector machine method. The modeling process proceeded until a highly accurate model was achieved, with the root mean square error percentage (RMSE%) being less than 40. The Latin hypercube approach was utilized for sample design, with 500 repetitions in this study. After selecting sampling points for each run using the Latin hypercube method, these points were mapped onto a detailed map, and the corresponding feature values were retrieved. The final map was created based on the extracted points. Subsequently, the latin hypercube approach was employed to generate soil property maps for each selected

https://doi.org/10.22067/jsw.2024.86936.1388

dataset. Validation was conducted using criteria such as the coefficient of explanation, root mean square error, and root mean square error in multiple iterations to ensure the accuracy of the generated maps.

Results and Discussion

The results distinctly illustrates the varied selection of sampling positions with each implementation of the Latin hypercube method. It is important to note that there may be some overlaps in different implementations. Consequently, the primary question arises: Is a one-time execution of the Latin hypercube sufficient for selecting study points? The findings indicate that the support vector machine model achieves satisfactory accuracy for all the examined characteristics. In the studied area, the environmental factors such as slope and elevation were identified as a significant predictors for estimating percentage of equivalent calcium carbonate.

Conclusion

In the present study, the accuracy of the latin hypercube method was assessed for selecting sampling location for digital soil mapping endeavors in Chaharmahal and Bakhtiari Province. Given the impracticality of collecting numerous field samples to evaluate the soil sampling method, this research aimed to employ simulation methods based on highly accurate maps for this purpose. The results indicate that the different outputs of the Latin hypercube method influence the accuracy of modeling, although this effect is also influenced by the specific feature under investigation and the extent of its variability within the study area. Considering that the Latin hypercube method is based on the principle that samples are randomly selected in each class of environmental parameters, it is suggested that future studies using this method should account for this principle. Adequate consideration should be given, and the selection of sampling locations should rely on multiple implementations of the Bhattacharya distance method to ensure robustness and reliability.

Keywords: Bhattacharyya distance, Digital soil mapping, Sampling position, Support vector machine

نشریه آب و خاک
https://jsw.um.ac.ir

مقاله پژوهشی
جلد ۳۸، شماره ۳، مرداد-شهریور ۱۴۰۳، ص. ۳۸۲-۳۶۷

ارزیابی دقت روش فرامکعب لاتین به‌منظور انتخاب موقعیت نقاط مطالعاتی برای تهیه نقشه‌ی رقومی ویژگی‌های خاک

زهره مصلح قهفرخی[1]*- ابوالفضل آزادی[2]

تاریخ دریافت: ۱۴۰۲/۱۱/۳۰
تاریخ پذیرش: ۱۴۰۳/۰۲/۱۸

چکیده

با توجه به اینکه دقت و صحت تمام اطلاعات خاک‌شناسی وابسته به بهترین گمانه‌زنی در مورد مکان تغییرات خاک‌ها در قالب تعیین الگوی نمونه‌برداری می‌باشد، انتخاب روشی کارآمد که بتواند به بهترین شکل این تغییرات را رصد نماید بسیار حائز اهمیت است. تاکنون مطالعات اندکی در رابطه با بررسی تأثیر تصادفی بودن انتخاب نمونه‌ها در روش فرامکعب لاتین بر صحت نقشه‌ها انجام شده است. این مطالعه با هدف ارزیابی دقت روش فرامکعب لاتین در انتخاب موقعیت نمونه‌برداری به‌منظور انجام مطالعات نقشه‌برداری رقومی خاک در منطقه‌ای از شهرستان بروجن در استان چهارمحال و بختیاری انجام شد. با توجه به اینکه، چندین مرتبه نمونه‌برداری میدانی برای ارزیابی روش نمونه‌برداری خاک امری غیرمنطقی است در این پژوهش تلاش گردید تا از روش‌های شبیه‌سازی بر اساس نقشه‌هایی با صحت بسیار بالا برای این منظور استفاده شود. فاصله باهاتاچاریا برای کمّی‌سازی فاصله بین توزیع احتمال جامعه اصلی و اجراهای مختلف روش فرامکعب لاتین استفاده گردید. نقشه ویژگی‌های خاک (درصد کربنات کلسیم معادل، رس و کربن آلی) عمق سطحی (صفر تا ۳۰ سانتی‌متر) با استفاده از روش ماشین‌بردار پشتیبان تهیه گردید و اعتبارسنجی شد. علاوه بر آن، انتخاب موقعیت نقاط نمونه‌برداری با استفاده از روش فرامکعب لاتین با تراکم ۲۰۰ نقطه با ۵۰۰ مرتبه اجرا انجام گردید. در هر مرحله، اعتبارسنجی برای پیش‌بینی ویژگی‌های خاک با استفاده از R^2، RMSE و RMSE% انجام شد. نتایج نشان داد که برای تمامی ویژگی‌های مورد بررسی، مدل ماشین‌بردار پشتیبان از صحت قابل قبولی (RMSE% کمتر از ۴۰) برخوردار می‌باشد. از سوی دیگر، نتایج گویای آن است که خروجی‌های مختلف روش فرامکعب لاتین در اجراهای مختلف آن بر صحت مدلسازی تأثیرگذار است و مقادیر RMSE مدل در حالت‌های مختلف برای درصد کربنات کلسیم معادل، رس و کربن آلی به‌ترتیب از ۱/۱، ۱/۱ و ۰/۰۲ تا ۳/۲، ۲ و ۰/۱۲ متغیر است. اگرچه این موضوع متأثر از ویژگی مورد بررسی و میزان تغییرات آن در منطقه مورد مطالعه نیز می‌باشد.

واژه‌های کلیدی: فاصله باهاتاچاریا، ماشین‌بردار پشتیبان، موقعیت نمونه، نقشه‌برداری رقومی

۱- استادیار پژوهش بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان چهارمحال و بختیاری، سازمان تحقیقات، آموزش و ترویج کشاورزی، شهرکرد، ایران
(*- نویسنده مسئول: Email: Z.mosleh@areeo.ac.ir)
۲- استادیار پژوهش، بخش تحقیقات خاک و آب، مرکز تحقیقات و آموزش کشاورزی و منابع طبیعی استان خوزستان، سازمان تحقیقات، آموزش و ترویج کشاورزی، اهواز، ایران
https://doi.org/10.22067/jsw.2024.86936.1388

مقدمه

تعیین دقیق ویژگی‌های خاک به‌دلیل تأثیرگذاری آن‌ها بر بسیاری از کارکردهای مهم خاک از جمله چرخه‌ی عناصر غذایی و همچنین به‌منظور مدیریت صحیح خاک بسیار حائز اهمیت است. داده‌های جمع آوری شده از خصوصیات خاک و اراضی طی مطالعات میدانی، منبع مهمی برای تهیه نقشه‌های خاک می‌باشند. اگرچه نمونه‌های برداشت شده به تعداد استاندارد یک مطالعه خاک‌شناسی در هر سطح دقت نمی توانند به‌طور کامل تمامی تغییرات خاک‌ها در منطقه را نشان دهند اما تنها منبع اطلاعاتی برای انجام هر پژوهشی هستند. از سوی دیگر، انتخاب مکان نمونه‌برداری یکی از مهم‌ترین مراحل مطالعات خاک شناسی است که با توجه به تغییرات مکانی خاک‌ها می‌تواند نقش مهمی در بیان ویژگی‌های خاک به‌همراه داشته باشد. نظر به محدودیت‌های مالی و زمانی، اتخاذ روش نمونه‌برداری مؤثر یکی از مهم‌ترین مسائل در هر پژوهش به‌ویژه مطالعات مربوط به کشاورزی و منابع طبیعی است. بنابراین، با توجه به اینکه صحت تمام داده‌ها و اطلاعات مطالعات خاک‌شناسی وابسته به بهترین گمانه‌زنی در مورد مکان تغییرات خاک‌ها در قالب تعیین الگوی نمونه‌برداری می‌باشد؛ انتخاب یک روش کارآمد که بتواند به بهترین شکل این تغییرات را رصد نماید بسیار حائز اهمیت است.

طراحی الگوی نمونه‌برداری از مراحل اصلی نقشه‌برداری خاک به شمار می‌رود و این طراحی به‌ویژه در مناطقی که توزیع مکانی خاک‌ها پیچیده است اهمیت بسیار زیادی در مطالعات خاک‌شناسی دارد (Yang *et al*., 2020; Brus, 2019). کیمبل و همکاران (Kimble *et al*., 2001) اظهار داشتند که تخمین صحیح ویژگی‌های خاک تا حد زیادی متأثر از روش نمونه‌برداری، تراکم نمونه‌برداری و تغییرپذیری ویژگی مورد نظر می‌باشد. به‌طور کلی میزان خطای موجود در مطالعات می‌تواند به دو گروه کلی شامل خطای نمونه‌برداری و خطای آزمایشی تفکیک گردند که میزان آن به‌ترتیب ۹۰ و ۱۰ درصد است (Lame & Defize, 1993). بنابراین، در بسیاری از موارد خطای ایجاد شده به‌دلیل طراحی نامناسب نمونه‌برداری به‌مراتب بسیار بیشتر از خطای ایجاد شده در هنگام آماده‌سازی نمونه، خطای دستگاه‌های اندازه‌گیری و آنالیز داده‌ها می‌باشد (Markert, 2007). از این رو، لزوم انتخاب یک طرح نمونه برداری مناسب بیش از پیش احساس می‌گردد. اگرچه باور عمومی بر این است که نمونه‌برداری خوب زمانی اتفاق می‌افتد که پراکنش جغرافیایی خوبی از نقاط نمونه‌برداری وجود داشته باشد اما در بعضی مواقع ممکن است علی‌رغم توزیع جغرافیایی مناسبی تغییرات موجود در منطقه به خوبی آشکار نشوند. تاکنون مطالعات مختلفی برای دستیابی به نمونه‌برداری بهتر، تکنیک‌های نمونه‌برداری را بر اساس داده‌های کمکی طراحی نموده‌اند. هنگل و همکاران (Hengl *et al*., 2003) بر اساس متغیرهای محیطی روشی را پیشنهاد کردند که واریانس کل در آن حداکثر باشد یعنی تغییرپذیری متغیر مورد مطالعه در نمونه برداشت شده حداکثر باشد.

روش فرامکعب لاتین[1] به‌عنوان روشی کارامد برای تعیین موقعیت نمونه‌برداری شناخته شده است و به‌طور گسترده‌ای در مطالعات استفاده می‌شود. در این روش، نمونه‌ها به‌طور تصادفی در هر طبقه از پارامترهای محیطی انتخاب می‌گردند. بنابراین، در هر مرحله از اجرای این روش، مجموعه نقاط انتخاب‌شده می‌تواند بسیار متفاوت باشد. از این‌رو، مجموعه نقاط مختلف می‌توانند بر صحت نقشه‌های تولیدی تأثیرگذار باشند. از سال ۲۰۰۶ که این روش توسط میناسنی و مک براتنی (Minasny & McBratney, 2006) پایه‌گذاری شد پژوهشگران مختلفی (Ramirez-Lopez *et al*., 2014; Adhikari *et al*., 2017; Zhang *et al*., 2022) این روش را طی مطالعات خود پیشنهاد داده‌اند. ژانگ و همکاران (Zhang *et al*., 2022) چهار روش نمونه‌برداری شامل شبکه‌ای، تصادفی شبکه‌ای و فرامکعب لاتین را مقایسه نمودند. آن‌ها اظهار داشتند که روش نمونه‌برداری شبکه‌ای و فرامکعب لاتین نتایج بهتری داشتند. روش شبکه‌ای پوشش بهتر و کامل‌تری از توزیع نمونه‌ها ارایه می‌دهد در حالی‌که روش فرامکعب لاتین فضای ویژگی بهتری را ارایه می‌دهد. از سوی دیگر، وادوکس و براس (Wadoux & Brus, 2021) اذعان داشتند که روش فرامکعب لاتین نسبت به سایر روش‌ها از خطای بیشتری برخوردار است و این موضوع زمانی قابل توجه می‌باشد که تراکم نمونه‌ها بیشتر است. ما و همکاران (Ma *et al*., 2020) روش فرامکعب لاتین را با روش نمونه برداری پوشش فضا[2] مقایسه نمودند و اظهار کردند که روش فرامکعب لاتین در بین چند تراکم مختلف ضعیف‌ترین صحت را ارایه داده است. خان و همکاران (Khan *et al*., 2023) بیان کردند که کارایی روش فرامکعب لاتین تا حدی زیادی به ویژگی مورد بررسی و روش مدلسازی وابسته است.

با این وجود تاکنون مطالعات اندکی در رابطه با بررسی تأثیر تصادفی بودن انتخاب نمونه‌ها در روش فرامکعب لاتین بر صحت نقشه‌ها انجام شده است. در روش فرامکعب لاتین، نمونه‌ها در هر طبقه از پارامترهای محیطی به‌صورت تصادفی انتخاب می‌گردند. بنابراین در هر مرتبه اجرای الگوریتم، نمونه‌های انتخاب‌شده می‌توانند متفاوت باشند. از این رو، صحت نقشه‌های تهیه‌شده در هر مرتبه می‌تواند کاملا متفاوت باشد. با این وجود، این که تا چه اندازه تصادفی بودن روش می

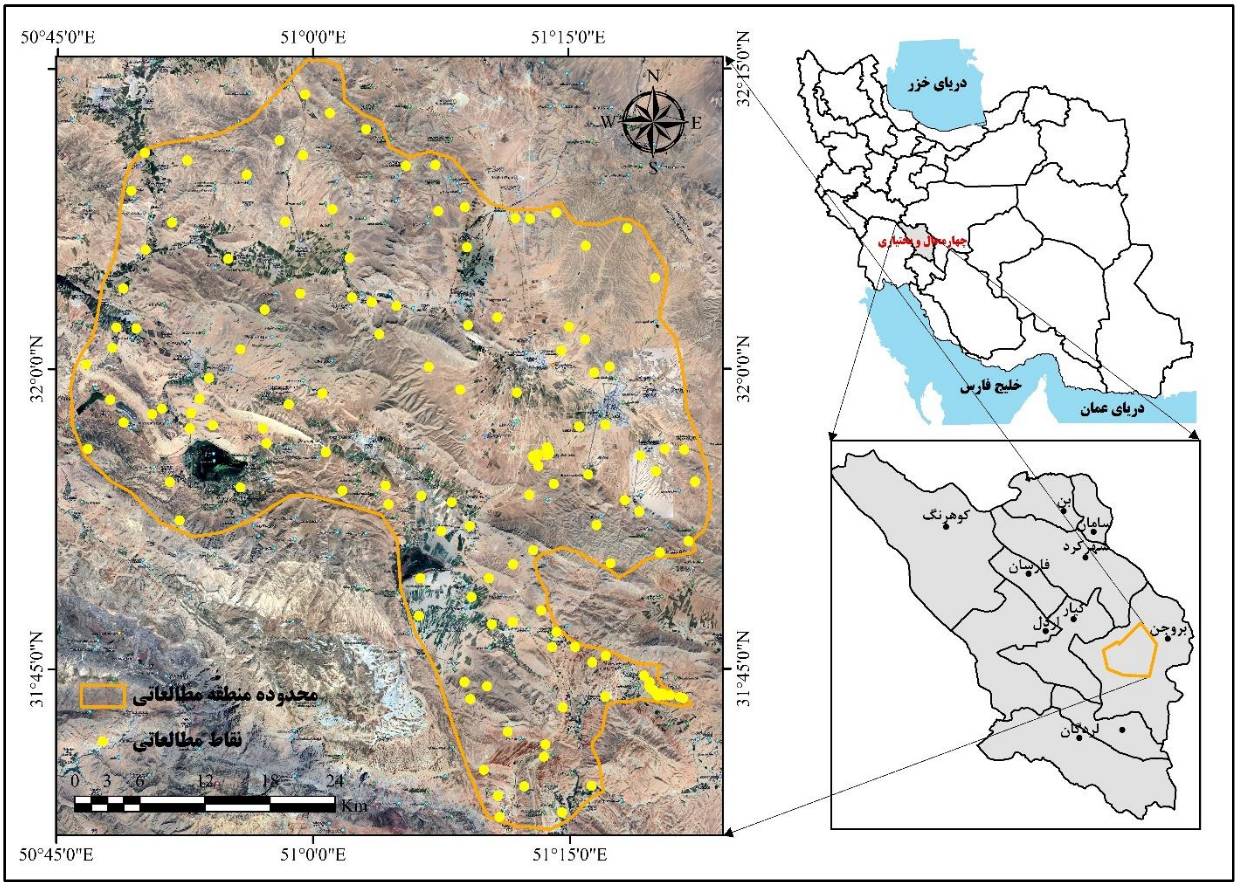
1- conditioned Latin Hypercube Sampling (cLHS)
2- Feature Space Coverage Sampling (FSCS)

تواند بر صحت نقشه‌های خروجی تأثیرگذار باشد به‌ندرت بررسی شده است. بنابراین، سوال اصلی این است که آیا روش نمونه‌برداری فرامکعب لاتین که در حال حاضر به‌طور گسترده‌ای استفاده می‌شود آیا همواره یک روش قابل اعتماد در این زمینه می‌باشد یا خیر؟ برتری روش فرامکعب لاتین نسبت به سایر روش‌های نمونه‌برداری به‌دلیل شباهت توزیع فراوانی این روش نمونه‌برداری به متغیرهای جامعه توسط پژوهشگران مختلفی از جمله (Brungard & Boettinger, 2010; Roudier *et al*., 2012; Thomas *et al*., 2012) بیان گردیده است. متاسفانه در هیچ یک از این مطالعات به واقعیت تغییرات خاک توجهی نشده است و پاسخ به این سوال هنوز در هاله‌ای از ابهام است که به چه دلیل روشی مانند فرامکعب لاتین به‌عنوان بهترین روش انتخاب گردیده است؟ معیار برتر بودن روش چیست و برترین روش نمونه برداری تا چه اندازه در بیان واقعیت خاک منطقه موفق می‌باشد؟ با توجه به اینکه چندین مرتبه نمونه‌برداری میدانی برای ارزیابی روش نمونه‌برداری خاک امری غیرمنطقی است در این پژوهش تلاش می‌شود تا از روش‌های شبیه‌سازی برای این منظور استفاده شود. هدف از پژوهش حاضر، ارزیابی دقت روش فرامکعب لاتین در تکرارهای مختلف به‌منظور تهیه نقشه ویژگی‌های خاک سطحی (درصد کربن آلی، رس و کربنات کلسیم معادل) در منطقه‌ای از شهرستان بروجن استان چهارمحال و بختیاری می‌باشد.

مواد و روش‌ها

توصیف منطقه‌ی مطالعاتی

منطقه‌ی مورد مطالعه با مساحت ۱۲ هزار هکتار بین طول‌های جغرافیایی "۳۵ '۴۷ °۵۰ تا "۲۹ '۲۳ °۵۱ شرقی و عرض‌های جغرافیایی "۳۷ '۳۶ °۳۱ تا "۴۸ '۱۵ °۳۲ شمالی و در شهرستان بروجن استان چهارمحال و بختیاری قرار گرفته است (شکل ۱). متوسط ارتفاع از سطح دریا و دمای سالیانه‌ی منطقه به‌ترتیب، ۲۳۳۸ متر و ۱۱/۵ درجه‌ی سانتی‌گراد و میزان بارندگی سالیانه آن ۲۵۰ میلی‌متر می‌باشد. منطقه عمدتا شامل تراس‌های آبرفتی جوان و قدیمی و سازند بختیاری می‌باشد.



شکل ۱- موقعیت محدوده و نقاط مطالعاتی

Figure 1- Location of the study area and sampling points

جمع‌آوری داده

برای انجام این پژوهش از داده‌های موروثی مطالعات خاک‌شناسی موجود در منطقه مورد مطالعه (Yekom Consulting Engineers. 1988) استفاده شد. برای این منظور مجموعه داده‌ای با تراکم ۲۵۰ نمونه در نظر گرفته شد. موقعیت مکانی نقاط نمونه‌برداری در شکل ۱ نشان داده شده است. ویژگی‌های مورد بررسی در این پژوهش، درصد کربنات کلسیم معادل، رس و کربن آلی خاک در عمق صفر تا ۳۰ سانتی‌متری می‌باشد.

تهیه پارامترهای محیطی

مدل رقومی ارتفاع آلوس پالسار[1] با قدرت تفکیک مکانی ۱۲/۵متر برای استخراج متغیرهای اجزای سرزمین به‌عنوان متغیرهای کمکی استفاده گردید. استخراج ویژگی‌های اولیه و ثانویه شامل درصد شیب[2]، جهت شیب[3]، انحنای خالص[4]، انحنای نیمرخ[5]، انحنای سطحی[6]، عمق دره[7]، شاخص زبری پستی و بلندی[8]، موقعیت نسبی شیب[9]، شاخص خیسی[10] و شاخص هموارى دره با درجه‌ی تفکیک بالا[11] با استفاده از نرم‌افزار SAGA-GIS انجام گردید.

مدلسازی با استفاده از ماشین‌بردار پشتیبان

در این مرحله، نقشه‌ی رقومی ویژگی‌های خاک شامل درصد کربنات کلسیم معادل، رس و کربن آلی خاک با اندازه پیکسل ۵۰ متر تهیه گردید. برای این منظور از داده‌های به‌دست آمده در مرحله جمع آوری داده و پارامترهای محیطی استفاده گردید و به‌منظور اجرای فرآیند مدل‌سازی از مدل ماشین‌بردار پشتیبان استفاده شد. این مدل، نخستین-بار توسط ولادیمر وپنیک[12] در سال ۱۹۶۵ ابداع شد و سپس توسط وپینک و کورینا کورتز[13] در سال ۱۹۹۵ برای جنبه غیرخطی پدیده‌ها ارایه گردید (We & Ai, 2008). هدف این مدل، یافتن بهترین مرز در بین داده‌ها می‌باشد؛ به‌گونه‌ای که بیشترین فاصله ممکن را از تمام دسته‌ها (بردارهای پشتیبان آن‌ها) داشته باشد. یک راه ساده برای انجام این کار و ساخت یک طبقه بهینه، محاسبه فاصله مرزهای به‌دست آمده با بردارهای پشتیبان هر طبقه و در نهایت انتخاب مرزی است که از طبقه‌های موجود، مجموعاً بیشترین فاصله را داشته باشد. این عمل تعیین مرز و انتخاب خط بهینه (در حالت کلی، ابرصفحه مرزی) به‌راحتی با انجام محاسبات ریاضی نه چندان پیچیده، قابل پیاده‌سازی است. لازم به ذکر است که در فرآیند مدل‌سازی، ابتدا تمامی پارامترهای محیطی به مدل داده شد و اهمیت نسبی این متغیرها در پیش‌بینی ویژگی‌های خاک با دستور varImp در نرم‌افزار R تعیین گردید. سپس، مؤثرترین متغیرها انتخاب شدند و مجدداً مدل‌سازی برای هر یک از ویژگی‌ها با توجه به اثرگذارترین متغیرها اجرا شد. در این مرحله، مدل‌سازی تا جایی انجام گردید که مدلی با صحت بالا (RMSE% کمتر از ۴۰) به‌دست آید (Hengl *et al*., 2004).

طراحی الگوی نمونه‌برداری

همان‌گونه که قبلا بیان گردید هدف اصلی این پژوهش بررسی میزان وابستگی صحت نقشه‌های تولیدی به مجموعه نمونه‌های انتخابی از روش فرامکعب لاتین می‌باشد. با توجه به اینکه بررسی تأثیر تصادفی بودن موقعیت نمونه بر صحت نقشه‌ها مستلزم انجام نمونه برداری در هر مرحله می‌باشد و این موضوع امری هزینه‌بر می‌باشد در این پژوهش از روش شبیه‌سازی استفاده گردید و نقشه ویژگی‌های خاک تهیه‌شده با استفاده از مدل ماشین‌بردار پشتیبان که دارای دقت قابل قبول بود به‌عنوان مجموعه داده قابل اطمینان در نظر گرفته شد (Yang *et al*., 2020). در این پژوهش، طراحی نمونه‌برداری با روش فرامکعب لاتین با ۵۰۰ تکرار انجام گردید. پس از انتخاب موقعیت نقاط نمونه‌برداری توسط روش فرامکعب لاتین در هر بار اجرا، این نقاط بر روی نقشه ویژگی‌های خاک که در مرحله قبل تهیه گردید و دارای صحت قابل قبول می‌باشد (RMSE% کمتر از ۴۰)؛ منتقل و مقادیر هر ویژگی استخراج شدند. سپس نقاط استخراج شده به‌عنوان مجموعه نقاط برای ایجاد نقشه‌ی مورد نظر مورد استفاده قرار گرفتند. در نهایت بر اساس هر یک از مجموعه داده‌های انتخاب‌شده توسط روش فرامکعب لاتین نقشه ویژگی‌های خاک تهیه گردید. به بیان دیگر، در هر مرتبه تکرار فرامکعب لاتین، مجوعه انتخاب شده به‌عنوان مجموعه داده منحصر به فرد برای فرآیند مدل‌سازی مورد استفاده قرار گرفت و با استفاده از روش ماشین‌بردار پشتیبان مدل‌سازی برای ویژگی‌های مختلف انجام شد. سپس، اعتبارسنجی در تکرارهای مختلف بر اساس معیارهای اعتبارسنجی از قبیل ضریب تبیین (رابطه ۱)، جذر میانگین مربع خطا[14] (رابطه ۲) و جذر میانگین مربع خطای نسبی[15] (رابطه ۳)

1- Alos Palsar
2- Slope
3- Aspect
4- Curvature
5- Plan curvature
6- Profile curvature
7- Valley depth
8- Terrain roughness index
9- Relative slope position
10- Wetness index
11- Multi-resolution valley bottom flatness index
12- Vladimir Vapnik
13- Corinna Cortes
14- Root Mean Squared Error (RMSE)
15- Normalized Root Mean Squared Error (RMSE %)

و بر اساس روابط زیر انجام گردید. لازم به ذکر است که مدل با استفاده از ۸۰ درصد داده‌ها آموزش داده شد و اعتبارسنجی آن با استفاده از ۲۰ درصد داده‌ها انجام گرفت. انتخاب داده‌های آموزشی و اعتبارسنجی به‌صورت تصادفی انجام شد. مدل‌سازی با استفاده از روش ماشین‌بردار پشتیبان با استفاده از بسته‌های caret و kernlab و الگوریتم فرامکعب لاتین با استفاده از بسته clhs در نرم‌افزار R 4.2.1 اجرا شد.

$$R^2 = \left[\frac{\sum_{i=1}^{n}(O_i - O_{ave})(P_i - P_{ave})}{\sqrt{\sum_{i=1}^{n}(O_i - O_{ave})^2 \, (P_i - P_{ave})^2}}\right]^2 \qquad (۱)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \qquad (۲)$$

$$\%\mathrm{RMSE} = \frac{\mathrm{RMSE}}{\mathrm{Oave}} \times 100 \qquad (۳)$$

در این روابط، Pi و Oi به‌ترتیب، مقادیر پیش‌بینی‌شده و واقعی، P_{ave} و O_{ave} میانگین مقادیر پیش‌بینی‌شده و واقعی، n تعداد مشاهدات می باشند. به‌طور کلی، مقادیر کوچکتر ضریب تبیین و مقادیر بزرگ‌تر جذر میانگین مربع خطا، نشان‌دهنده‌ی خطای بیش‌تر در تخمین می‌باشند.

فاصله‌ی باهاتاچاریا[1]

فاصله‌ی باهاتاچاریا، یک اندازه‌گیری آماری برای کمّی‌سازی فاصله‌ی میان دو توزیع احتمال می‌باشد (Khan *et al*., 2023). با توجه به اینکه، این روش نمونه‌برداری از توزیع پارامترهای محیطی پیروی می‌کند و دنباله‌های توزیع از شانس یکسانی نسبت به قسمت‌های مرکزی آن برخوردار می‌باشند؛ پس از آنکه مدل‌سازی در تکرارهای مختلف روش فرامکعب لاتین انجام شد، فاصله باهاتاچاریا برای تکرارهای مختلف و به ازای پارامترهای محیطی مختلف محاسبه گردید (Khan *et al*., 2023). منظور از این فاصله، میزان مشابهت بین دو توزیع احتمال می‌باشد. این فاصله مطابق با روابط زیر تعیین گردید (Khan *et al*., 2023):

$$\omega_{B\,(a,b)} = -\ln(\theta_C\ (a,b)) \qquad (٤)$$

$$\theta_C\,(a,b) = \int \sqrt{a\,(x) b\,(x)}.\,dx \qquad (٥)$$

در این رابطه، ω_B فاصله باهاتاچاریا بین دو توزیع احتمال a و b و θ_D ضریب باهاتاچاریا می‌باشد (روابط ۴ و ۵). توزیع احتمال‌های a و b مربوط به جامعه‌های مورد هدف می‌باشد. در این پژوهش، توزیع احتمال a مربوط به پارامترهای محیطی کل منطقه و توزیع احتمال b مربوط به پارامترهای محیطی نقاط نمونه‌برداری حاصل از روش فرامکعب لاتین در هر تکرار می‌باشد.

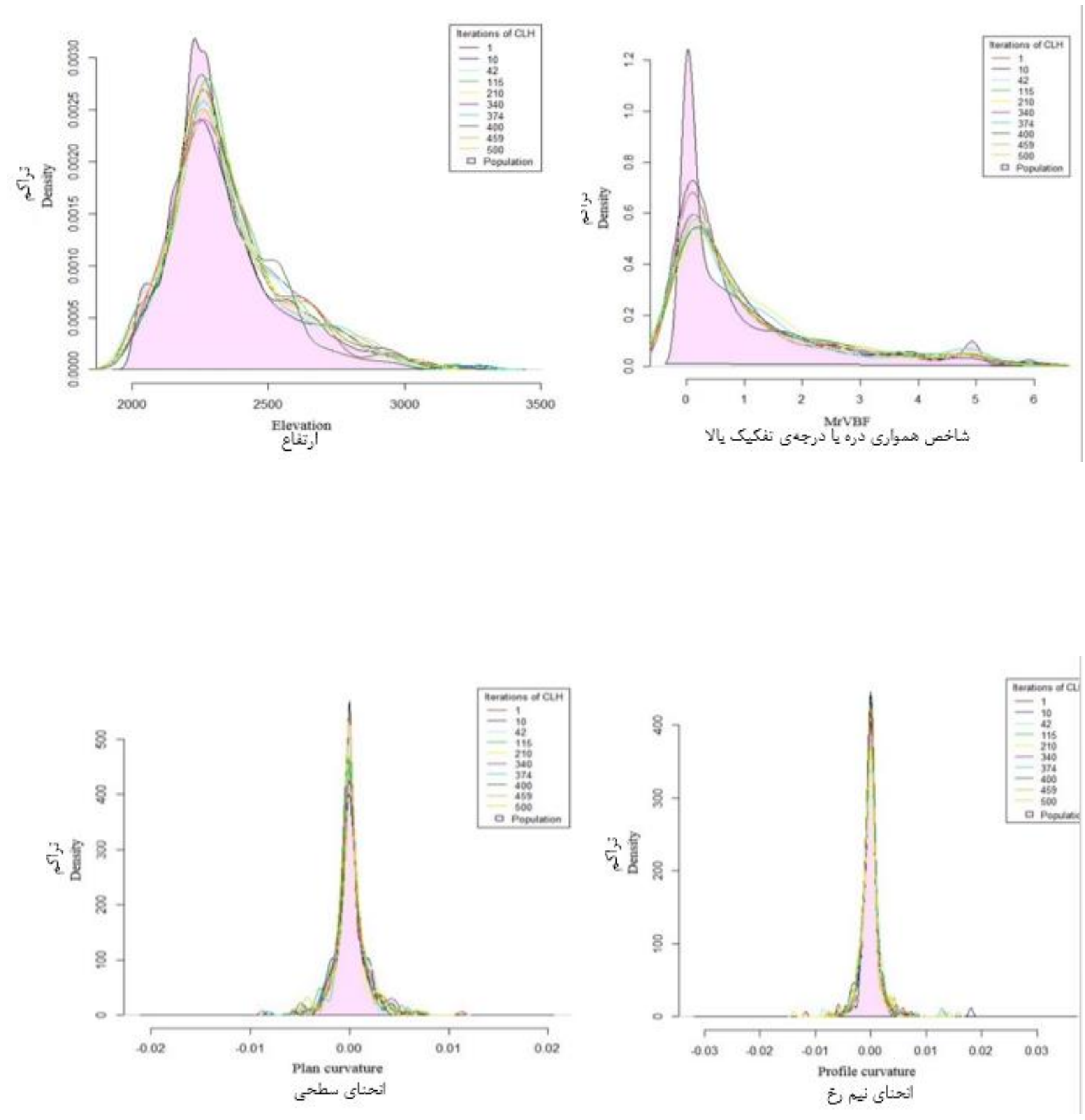
نتایج و بحث

نمودار توزیع احتمال برخی از پارامترهای محیطی برای جامعه اصلی و مقادیر آن‌ها در تکرارهای مختلف روش فرامکعب لاتین در شکل ۲ ارایه شده است. نمودارها حاکی از آن است که توزیع احتمال ورودی‌ها دارای چولگی می‌باشد. همچنین، مقایسه‌ی مقادیر پارامترهای محیطی در هر یک از حالت‌های اجرای روش فرا مکعب لاتین، اختلافاتی را با جامعه اصلی نشان می‌دهد. در رابطه با برخی از پارامترها مانند ارتفاع، انحنای سطحی و انحنای نیمرخ این اختلافات بسیار اندک است اما در مورد برخی از پارامترهای محیطی ازجمله شاخص همواری دره با درجه‌ی تفکیک بالا و موقعیت نسبی شیب تفاوت‌ها چشمگیرتر می‌باشد. بنابراین، با توجه به اینکه توزیع احتمال پارامترهای محیطی در منطقه مورد بررسی بسیار متفاوت است و یکی از ویژگی‌های اساسی روش فرامکعب لاتین انتخاب تصادفی نمونه‌ها از هر گروه می‌باشد می توان انتظار داشت که این موضوع بر روی خروجی مدل‌سازی تأثیرگذار باشد و در هر مرتبه اجرای الگوریتم، نمونه‌های مختلفی انتخاب گردند.

پراکنش نقاط انتخابی توسط روش فرامکعب لاتین در تکرارهای مختلف در شکل ۳ ارایه شده است. در این شکل، پراکنش نقاط برای ۱۰ تکرار از ۵۰۰ مرتبه تکرار نشان داده شده است. همانگونه که این شکل به وضوح نشان می‌دهد در هر مرتبه اجرای روش فرامکعب لاتین موقعیت‌های مختلف نمونه‌برداری انتخاب می‌گردند. اگرچه ذکر این نکته ضروری می‌باشد که ممکن است در بعضی از نقاط در اجراهای مختلف هم‌پوشانی نیز وجود داشته باشد.

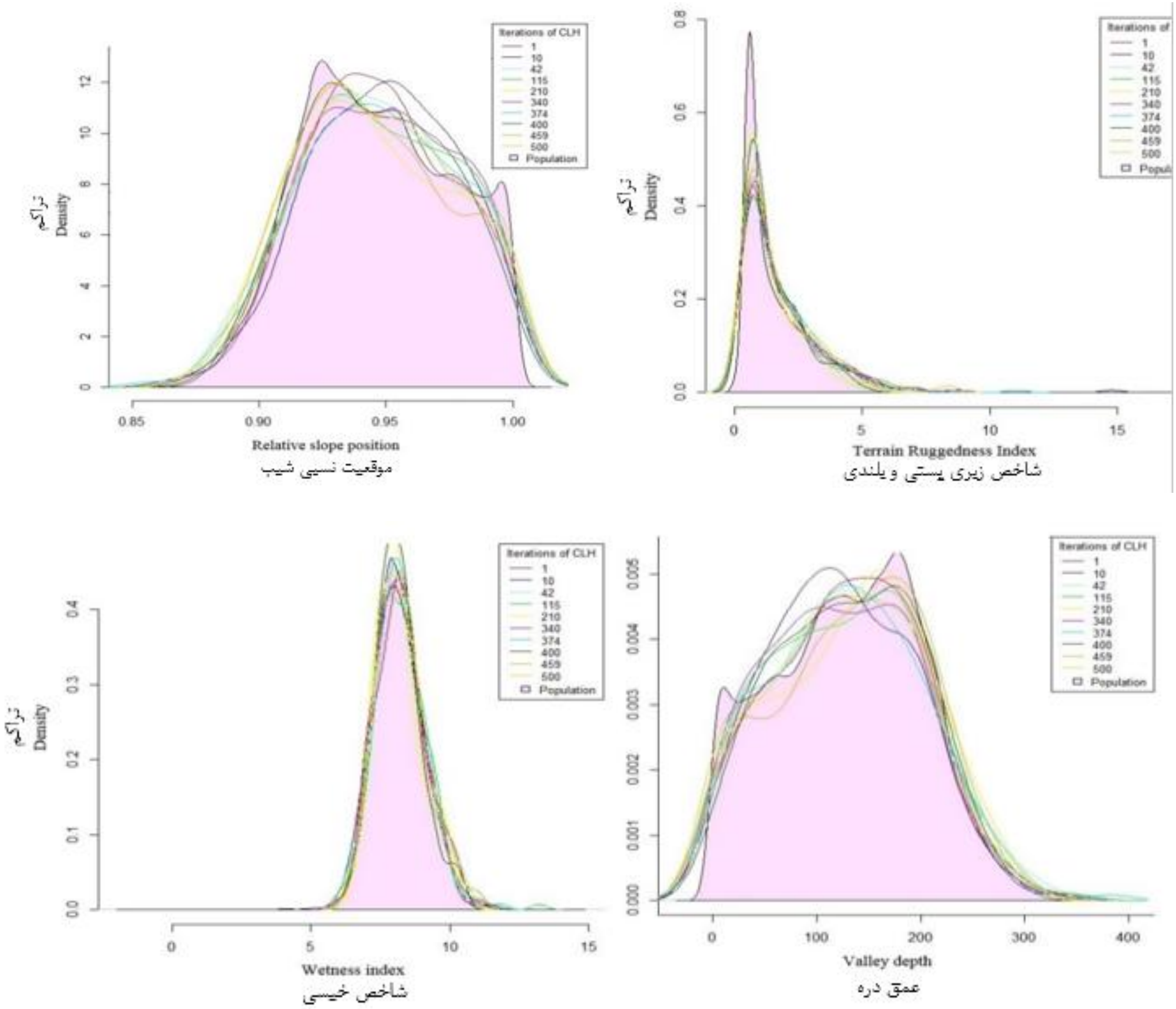
نتایج اعتبارسنجی پیش‌بینی ویژگی‌های خاک با استفاده از مدل ماشین‌بردار پشتیبان در جدول ۱ ارایه شده است. نتایج نشان می‌دهد که برای تمامی ویژگی‌های مورد بررسی، مدل ماشین‌بردار پشتیبان از صحت قابل قبولی برخوردار بوده است. در منطقه مورد مطالعه، بر اساس مقادیر اهمیت نسبی، پارامترهای محیطی شیب و ارتفاع برای پیش‌بینی درصد کربنات کلسیم معادل به‌عنوان مهم‌ترین پارامترها انتخاب شدند. همچنین، شاخص همواری دره با درجه تفکیک بالا، عمق دره و شاخص موقعیت نسبی شیب توسط مدل به‌عنوان مهم‌ترین پارامترها برای پیش‌بینی درصد رس خاک انتخاب شدند. پارامترهای عمق دره و شاخص خیسی مهم‌ترین تأثیر را در پیش‌بینی کربن آلی خاک داشته‌اند (جدول ۱).

1- Bhattacharyya distance



شکل ۲- نمودار توزیع احتمال برخی از پارامترهای محیطی برای جامعه اصلی و اجراهای مختلف روش فرامکعب لاتین

Figure 2- Probability density functions for some auxiliary information for main population and different runs of conditioned latin



ادامه شکل ۲- نمودار توزیع احتمال برخی از پارامترهای محیطی برای جامعه اصلی و اجراهای مختلف روش فرامکعب لاتین

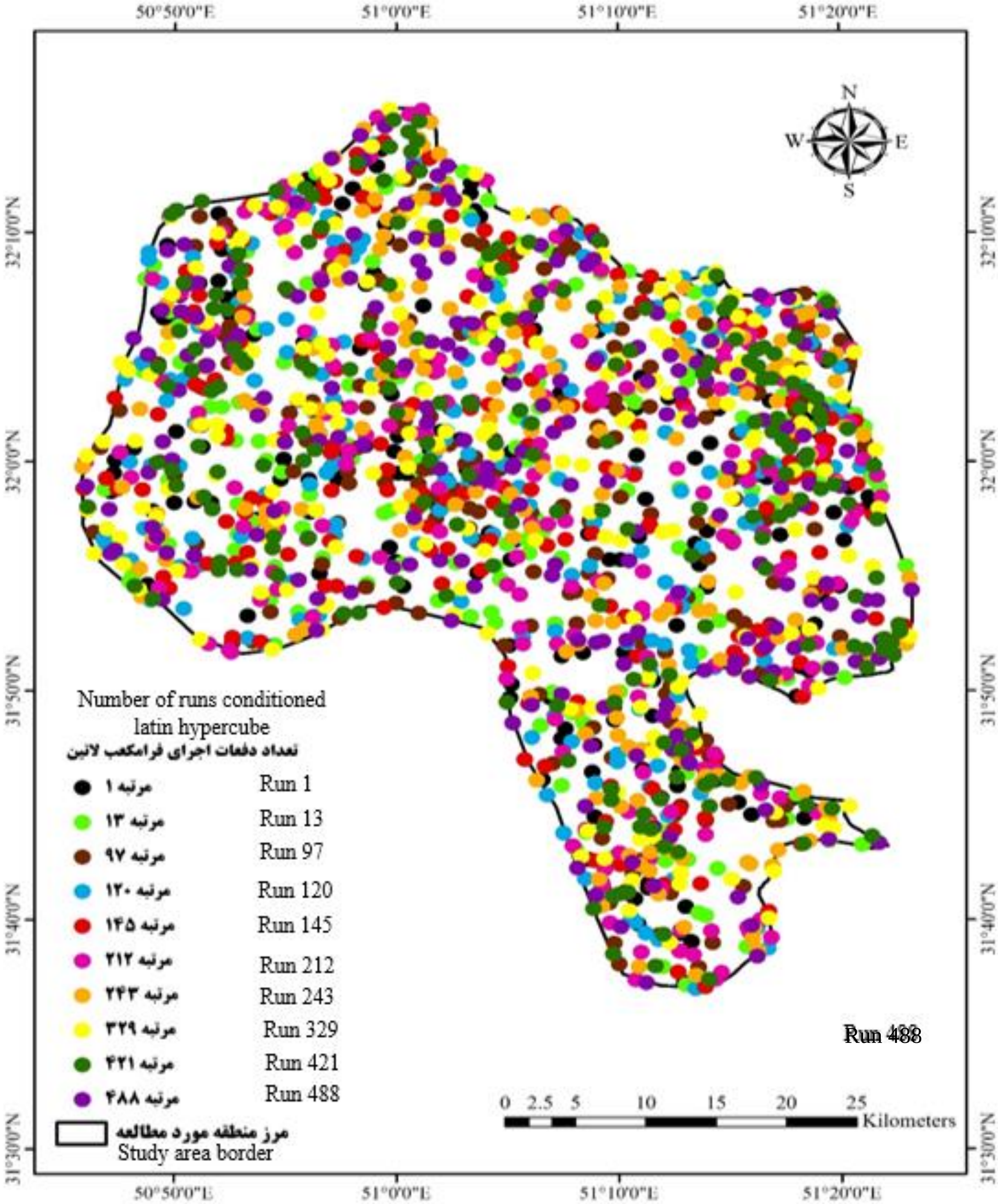
Continued Figure 2- Probability density functions for some auxiliary information for main population and different runs of conditioned latin hypercube

جدول ۱- نتایج مدل‌سازی ویژگی‌های خاک با استفاده از مدل ماشین‌بردار پشتیبان

Table 1- Results of modelling soil properties using support vector machine

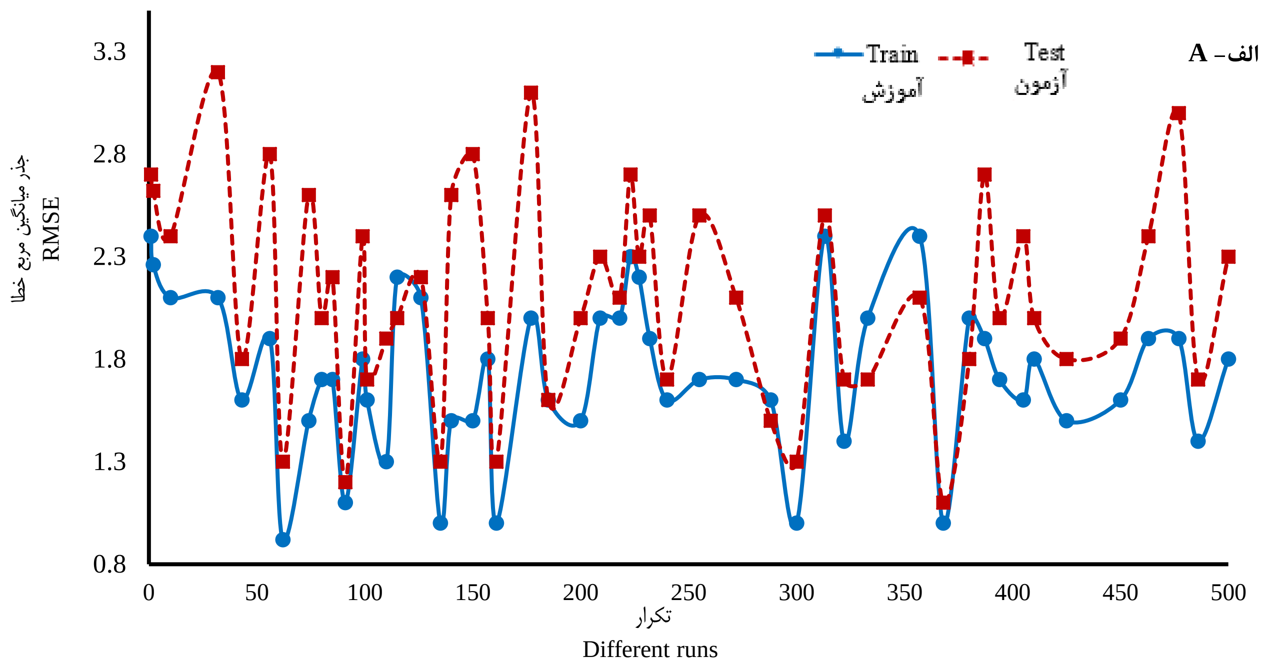
ویژگی خاک **Soil properties**	پارامترهای محیطی انتخاب شده **Selected auxiliary data**	نوع داده **Data**	**R^2**	**ME**	**RMSE**	**%RMSE**
درصد کربنات کلسیم معادل Calcium carbonate equivalent	شیب، ارتفاع Slope, Elevation	آموزش Training	0.84	1.43	6.23	21.45
		آزمون Testing	0.82	1.80	10.50	35.12
رس Clay	شاخص هموارى دره با درجه‌ی تفکیک بالا، عمق دره، شاخص موقعیت نسبی شیب Multi-resolution valley bottom flatness index, Relative slope position	آموزش Training	0.87	2.00	9.97	32.52
		آزمون Testing	0.80	1.00	10.72	37.14
کربن آلی Organic Carbon	عمق دره، شاخص خیسی Valley depth, Wetness index	آموزش Training	0.86	0.05	0.28	26.30
		آزمون Testing	0.78	0.01	0.37	36.23

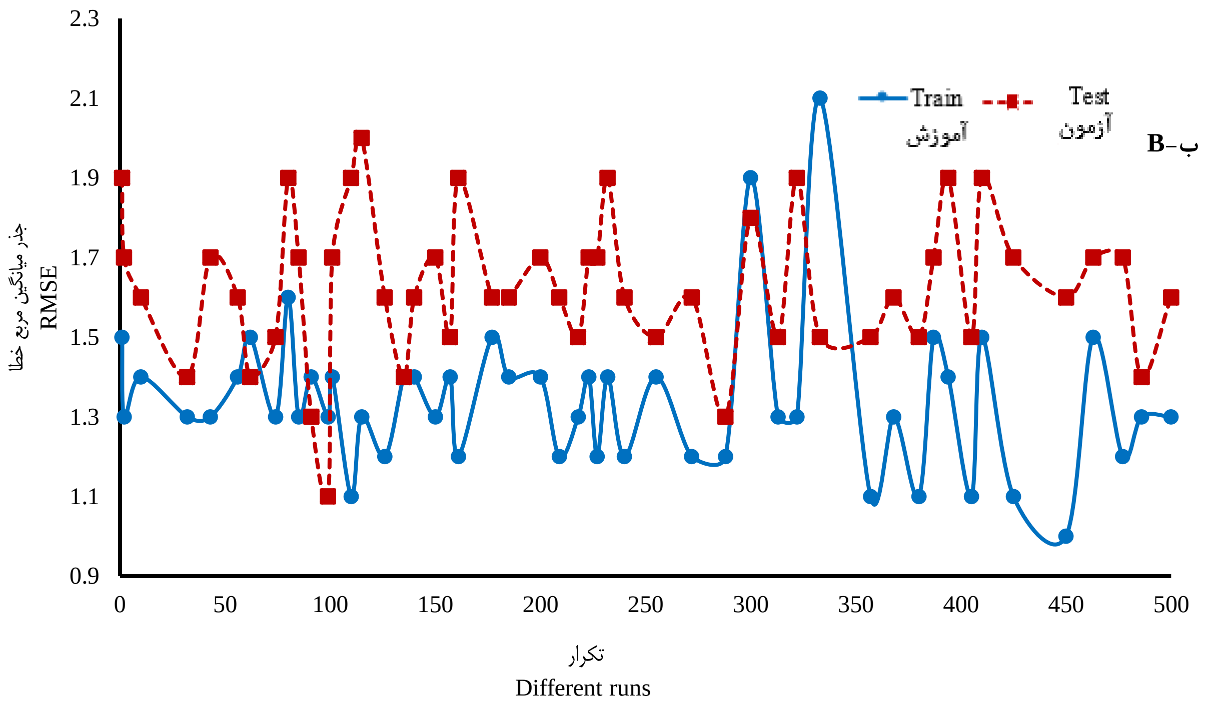
شکل ۴ مقادیر RMSE مدل را در تکرارهای مختلف روش فرامکعب لاتین برای پیش‌بینی ویژگی‌های خاک مورد برررسی نشان می‌دهد. لازم به ذکر است با توجه به اینکه امکان نمایش نتایج ۵۰۰ تکرار در مقاله وجود نداشت تنها نتایج مربوط به ۵۰ تکرار ارایه شده است. نتایج گویای آن است که خروجی‌های مختلف روش فرامکعب لاتین بر صحت مدل‌سازی تأثیرگذار است اگرچه این موضوع متأثر از ویژگی مورد بررسی و میزان تغییرات آن در منطقه مورد مطالعه نیز می‌باشد. نتایج نشان می‌دهد که در رابطه با ویژگی درصد کربنات کلسیم معادل مقادیر RMSE از ۱/۱ تا ۳/۲ در نوسان است و ضریب تغییرات برای تکرارهای مختلف ۲۴/۱ درصد می‌باشد. در مورد درصد رس خاک، مقادیر حداقل و حداکثر RMSE به‌ترتیب ۱/۱ و ۲ می‌باشد و ضریب تغییرات ۱۱/۴ درصد است. پیش‌بینی مقادیر کربن آلی خاک در تکرارهای مختلف حاکی از آن است که مقادیر RMSE از ۰/۰۲ تا ۰/۱۲ در نوسان است و ضریب تغییرات برای تکرارهای مختلف ۳۴/۱ درصد می‌باشد.



شکل ۳- پراکنش نقاط انتخابی توسط روش فرامکعب لاتین در اجراهای مختلف (۱۰ مرتبه تکرار از ۵۰۰ مرتبه به‌عنوان نمونه)

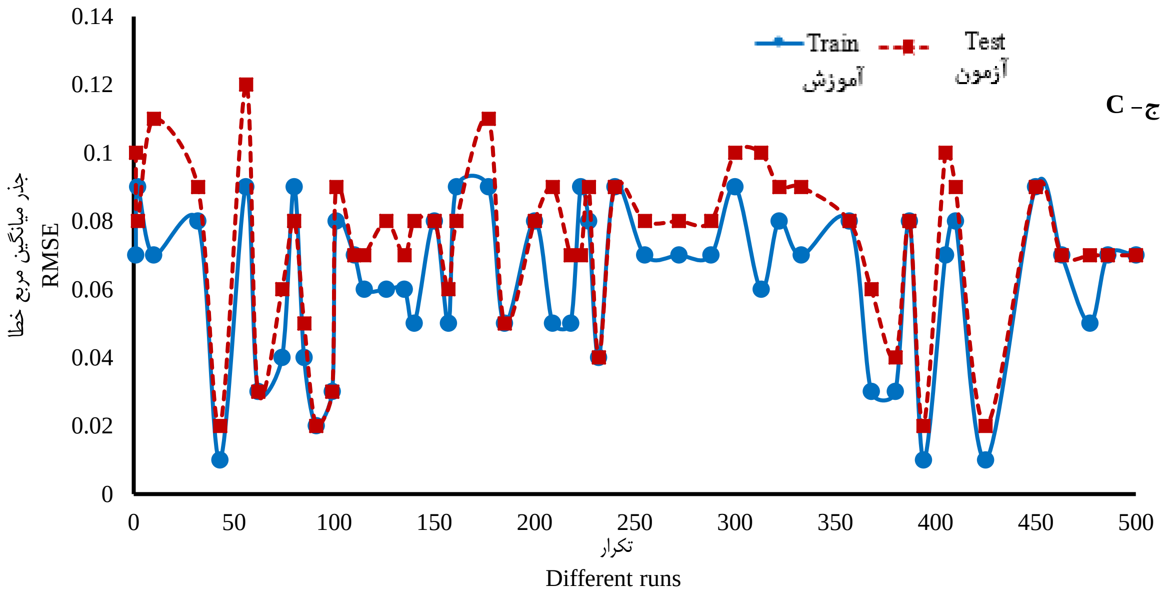
Figure 3- Spatial variations of selected sampling points using conditioned latin hypercube at different runs (10 times from 500 as an example)





شکل ٤- تغییرات مقادیر RMSE در تکرارهای مختلف روش فرامکعب لاتین برای پیش‌بینی (الف): درصد کربنات کلسیم معادل (ب) رس و (ج) کربن آلی خاک

Figure 4- Variation of RMSE at different runs for predicting calcium carbonate equivalent (A), clay (B) and organic carbon (C)

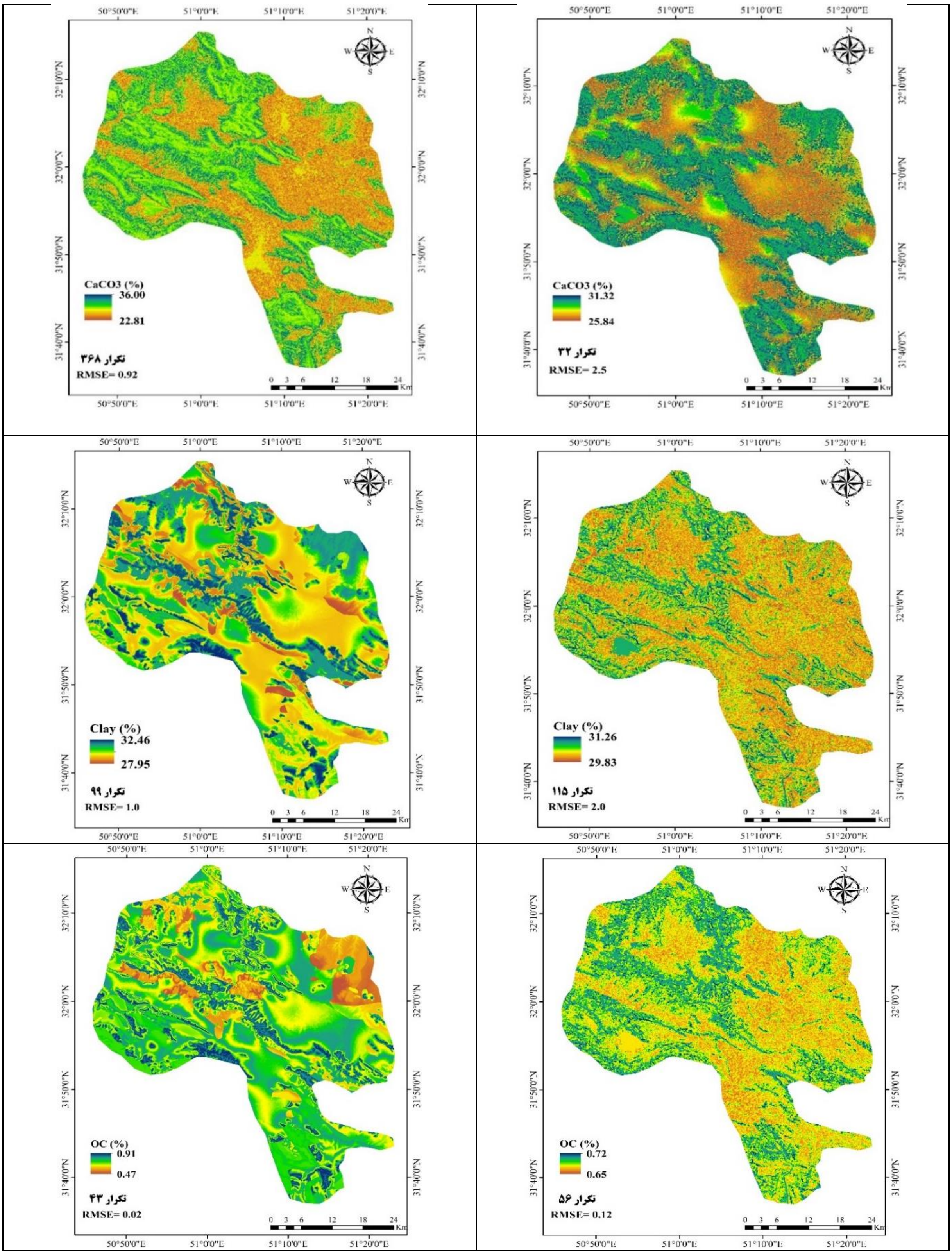


ادامه شکل ٤- تغییرات مقادیر RMSE در تکرارهای مختلف روش فرامکعب لاتین برای پیش‌بینی (الف): درصد کربنات کلسیم معادل (ب) رس و (ج) کربن آلی خاک

Continued Figure 4- Variation of RMSE at different runs for predicting calcium carbonate equivalent (A), clay (B) and organic carbon (C)

نقشه‌های رقومی ویژگی‌های خاک شامل درصد کربنات کلسیم معادل، درصد رس و کربن آلی با استفاده از مدل ماشین‌بردار پشتیبان برای نمونه‌های حاصل از اجراهای مختلف روش فرامکعب لاتین در شکل ۵ آورده شده است. با توجه به محدودیت ارایه تمامی نقشه‌ها، تلاش گردید تا برای هر ویژگی نقشه بهترین حالت (کمترین خطا) و بدترین حالت (بیشترین خطا) آورده شود. نقشه‌های به‌دست آمده به خوبی گویای آن هستند که خروجی نهایی مدل‌سازی که نقشه‌ها می باشد متأثر از روش اجرای الگوریتم فرامکعب لاتین می‌باشد. همانگونه که مشاهده می‌گردد نقشه‌های تهیه‌شده بر اساس اجراهای مختلف الگوریتم فرامکعب لاتین از نظر مقادیر هر یک از ویژگی‌ها و روند تغییرات آن ویژگی در منطقه تفاوت‌های قابل توجهی دارند. بنابراین، می‌توان به این سوال پاسخ داد که روش فرامکعب لاتین همیشه روش ایده آل برای انتخاب مکان نمونه‌برداری نیست و نمی‌توان به صرف یک مرتبه انجام آن اطمینان حاصل نمود که بهترین پراکنش نمونه-برداری وجود دارد و تغییرات منطقه به‌طور کامل قابل شناسایی است. از این رو، به نظر می‌رسد که اگر این روش به‌عنوان مبنایی برای نمونه‌برداری قرار می‌گیرد حتما باید به‌طور کامل بررسی گردد و بهترین حالت برای اجرای این روش را انتخاب نمود.

جدول ۲ مقادیر فاصله باهاتاچاریا را به ازای تکرارهای مختلف و برای پارامترهای محیطی که تغییرات آن‌ها چشمگیر بود را به همراه مقادیر p-value آن‌ها نشان می‌دهد. زمانی که مقادیر p-value کوچکتر از ۰/۰۵ باشد حاکی از آن است که احتمال آنکه فاصله میان دو توزیع اتفاقی باشد بسیار اندک است و می‌توان نتیجه گرفت که دو توزیع متفاوت از یکدیگر هستند (Bhattacharyya, 1943). خان و همکاران (Khan *et al*., 2023) اظهار داشتند که با افزایش تراکم نمونه‌برداری، میانگین نمونه‌ها به میانگین جامعه اصلی نزدیک می‌شود و از این رو مقادیر فاصله باهاتاچاریا کاهش می‌یابد. در جدول ۲ تکرارهایی که مقادیر p-value آن‌ها کوچکتر از ۰/۰۵ می‌باشد با رنگ آبی نشان داده شده است. بر اساس نتایج این جدول می‌توان بیان نمود که در برخی از اجراهای الگوریتم فرامکعب لاتین، اختلاف پارامترهای محیطی بین جامعه اصلی و نقاط انتخابی اندک و در برخی موارد این فاصله بیشتر است. علاوه بر آن، مشاهده می‌گردد که در یک اجرای به خصوص می‌تواند پارامترهای محیطی مختلفی برای ایجاد تفاوت در دو جامعه وجود داشته باشد. برای مثال، در تکرار ۱۰ مشاهده می-گردد که از بین پارامترهای انتخابی تنها شاخص زبری پستی و بلندی اختلاف آن چشمگیر است در حالی‌که در تکرار ۲۰۰ این اختلاف برای ارتفاع و عمق دره مشاهده می‌شود و در نهایت در برخی از تکرارها برای اکثر پارامترهای محیطی این اختلاف مشاهده می‌گردد برای نمونه در تکرار ۵۰۰، سه پارامتر محیطی ارتفاع، شاخص زبری پستی و بلندی و عمق دره در ایجاد اختلاف تأثیرگذار می‌باشند (جدول ۲).



شکل ۵- نقشه ویژگی‌های خاک با صحت‌های مختلف حاصل از اجراهای مختلف روش فرامکعب لاتین با استفاده از مدل ماشین‌بردار پشتیبان

Figure 5- Soil properties maps with different accuracy from different runs of conditioned latin hypercube sampling using support vector machine

جدول ۲- مقادیر فاصله باهاتاچاریا (d) و p-value برای تکرارهای مختلف اجرای الگوریتم فرامکعب لاتین به ازای پارامترهای محیطی مختلف

Table 2- The Bhattacharyya distance and p-value for different runs of Conditioned Latin Hypercube for different auxiliary information

شماره تکرار Number of run	پارامترهای محیطی Auxiliary information						شماره تکرار N. run	پارامترهای محیطی Auxiliary information					
	ارتفاع Elevation		شاخص زبری پستی و بلندی Terrain roughness index		عمق دره Valley depth			ارتفاع Elevation		شاخص زبری پستی و بلندی Terrain roughness index		عمق دره Valley depth	
	d	value-p	d	value-p	d	value-p		d	value-p	d	value-p	d	value-p
1	0.07	0.62	0.07	0.71	0.08	0.54	218	0.09	0.32	0.11	0.14	0.06	0.79
2	0.08	0.54	0.08	0.46	0.08	0.46	223	0.06	0.86	0.07	0.71	0.09	0.39
10	0.06	0.79	0.14	0.03	0.09	0.39	227	0.06	0.86	0.06	0.86	0.07	0.7
32	0.16	0.011	0.07	0.71	0.08	0.46	232	0.05	0.92	0.08	0.46	0.1	0.27
43	0.19	0.001	0.08	0.54	0.05	0.96	240	0.06	0.79	0.06	0.86	0.04	0.99
56	0.14	0.03	0.08	0.54	0.15	0.01	255	0.12	0.11	0.06	0.86	0.15	0.02
62	0.1	0.22	0.11	0.14	0.08	0.46	272	0.08	0.46	0.07	0.62	0.1	0.22
74	0.08	0.46	0.21	0.0002	0.09	0.39	288	0.06	0.79	0.07	0.62	0.07	0.7
80	0.11	0.14	0.09	0.39	0.09	0.32	300	0.11	0.17	0.12	0.11	0.11	0.17
85	0.08	0.46	0.15	0.01	0.07	0.71	313	0.07	0.62	0.11	0.17	0.1	0.22
91	0.09	0.32	0.21	0.0002	0.1	0.27	322	0.07	0.62	0.07	0.72	0.05	0.96
99	0.06	0.86	0.11	0.14	0.08	0.46	333	0.09	0.32	0.08	0.54	0.08	0.46
101	0.07	0.71	0.08	0.46	0.07	0.7	357	0.12	0.08	0.06	0.79	0.06	0.79
110	0.08	0.54	0.09	0.39	0.08	0.46	368	0.1	0.22	0.19	0.001	0.04	0.99
115	0.08	0.54	0.11	0.17	0.11	0.17	380	0.13	0.05	0.06	0.86	0.05	0.96
126	0.06	0.86	0.11	0.14	0.1	0.27	387	0.08	0.46	0.05	0.96	0.1	0.27
135	0.08	0.46	0.11	0.17	0.11	0.17	394	0.13	0.06	0.13	0.06	0.08	0.46
140	0.08	0.54	0.12	0.08	0.06	0.86	405	0.09	0.39	0.11	0.17	0.1	0.27
150	0.13	0.05	0.08	0.46	0.06	0.86	410	0.1	0.27	0.1	0.27	0.09	0.32
157	0.1	0.27	0.11	0.17	0.1	0.27	425	0.11	0.17	0.1	0.22	0.08	0.46
161	0.08	0.54	0.08	0.46	0.12	0.08	450	0.16	0.008	0.1	0.27	0.08	0.46
177	0.14	0.02	0.11	0.17	0.05	0.92	463	0.09	0.39	0.1	0.27	0.07	0.62
185	0.06	0.79	0.07	0.62	0.05	0.92	477	0.07	0.72	0.16	0.008	0.1	0.22
200	0.17	0.006	0.1	0.27	0.14	0.03	486	0.06	0.79	0.07	0.62	0.06	0.79
209	0.08	0.54	0.09	0.32	0.07	0.7	500	0.15	0.02	0.13	0.05	0.13	0.05

نتیجه‌گیری

در پژوهش حاضر، دقت روش فرامکعب لاتین در انتخاب موقعیت نمونه‌برداری به‌منظور انجام مطالعات نقشه‌برداری رقومی خاک در منطقه‌ای از شهرستان بروجن استان چهارمحال و بختیاری مورد ارزیابی قرار گرفت. با توجه به اینکه، چندین مرتبه نمونه‌برداری میدانی برای ارزیابی روش نمونه‌برداری خاک امری غیرمنطقی است در این پژوهش تلاش گردید تا از روش‌های شبیه‌سازی بر اساس نقشه‌هایی با دقت بسیار بالا برای این منظور استفاده شود. نتایج گویای آن است که خروجی‌های مختلف روش فرامکعب لاتین بر صحت مدلسازی تأثیرگذار است اگرچه این موضوع متأثر از ویژگی مورد بررسی و میزان تغییرات آن در منطقه مورد مطالعه نیز می‌باشد. با توجه به اینکه روش فرامکعب لاتین بر این اصل استوار است که نمونه‌ها به‌طور تصادفی در هر طبقه از پارامترهای محیطی انتخاب می‌گردند پیشنهاد می‌گردد که در مطالعات آتی که قرار است روش فرامکعب لاتین استفاده گردد حتما به این موضوع توجه کافی شود و بعد از چندین اجرای روش فاصله باهاتاچاریا محاسبه و بر آن اساس تصمیم به انتخاب مکان نمونه‌برداری شود.

References

1. Adhikari, K., & Hartemink, A.E. (2017). Soil organic carbon increases under intensive agriculture in the Central Sands, Wisconsin, USA. *Geoderma Regional*, *10*(1), 115–125. https://doi.org/10.1016/j.geodrs.2017.07.003
2. Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–109.
3. Brungard, C.W., & Boettinger, J.L. (2010). *Conditioned latin hypercube sampling: optimal sample size for digital soil mapping of arid rangelands in Utah, USA*. p. 67-75. In: Digital Soil Mapping. Springer, Netherlands.
4. Brus, D.J. (2019). Sampling for digital soil mapping: A tutorial supported by R scripts. *Geoderma*, *338*(1), 464–480.

https://doi.org/10.1016/j.geoderma.2018.07.036

5. Hengl, T., Huvelink, G.B.M., & Stein, A. (2004). A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma*, *120*, 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018
6. Hengl, T., Rossiter, D.G., & Stein, A. (2003). Soil sampling strategies for spatial prediction by correlation with auxiliary maps. *Australian Journal of Soil Research*, *41*(8), 1403-1422. https://doi.org/10.1071/SR03005
7. Khan, A., Aitkenhead, M., Stark, C.R., & Ehsan Jorat, M. (2023). Optimal sampling using Conditioned Latin Hypercube for digital soil mapping: An approach using Bhattacharyya distance. *Geoderma*, *439*(1), 116660. https://doi.org/10.1016/j.geoderma.2023.116660
8. Kimble, J.M., Grossman, R.B., & Samson-Liebig, S.E. (2001). *Methodology for sampling and preparation for soil carbon determination*. p. 15–30. In: Lal, R., J.M. Kimble, R.F. Follett, and B.A. Stewart (eds.) Assessment methods for soil carbon, Lewis Publishers, Boca Raton.
9. Lame, F.P.J., & Defize, P.R. (1993). Sampling of contaminated soil: Sampling error in relation to sample size and segregation. *Environmental Science and Technology*, *27*(10), 2035–2044.
10. Ma, T., Brus, D.J., Zhu, A.-X., Zhang, L., & Scholten, T. (2020). Comparison of conditioned Latin hypercube and feature space coverage sampling for predicting soil classes using simulation from soil maps. *Geoderma*, *370*(1), 114366. https://doi.org/10.1016/j.geoderma.2020.114366
11. Markert, B. (2007). *Quality assurance of plant sampling and storage.* p. 215-254. In: Quevauviller P (ed.) Qualilty Assurance in Environmental Monitoring: Sampling and Sample Preatreatment. Wiley-VCH, Verlag GmbH, Weinheim.
12. Minasny, B., & McBratney, A.B. (2006). A Conditioned Latin Hypercube method for sampling in the presence of ancillary information. *Computers & Geosciences*, *32*(9), 1378–1388. https://doi.org/10.1016/j.cageo.2005.12.009
13. Ramirez-Lopez, L., Schmidt, K., Behrens, T., van Wesemael, B., Dematte, J.A., & Scholten, T. (2014). Sampling optimal calibration sets in soil infrared spectroscopy. *Geoderma*, *226*(1), 140–150. https://doi.org/10.1016/j.geoderma.2014.02.002
14. Roudier, P., Beaudette, D.E., & Hewitt, A. (2012). *A conditioned Latin hypercube sampling algorithm incorporating operational constraints*. p. 227-231. In Minasny B, Malone B P, McBratney A B (eds.) Digital Soil Assessments and Beyond. CRC Press, London.
15. Thomas, M., Odgers, N.P., Ringrose-Voase, A., Grealish, G., Glover, M., & Dowling, T. (2012). *Soil survey design for management-scale digital soil mapping in a mountainous southern Philippine catchment*. p. 233-238. In Minasny B, Malone B P, McBratney A B (eds.) Digital Soil Assessments and Beyond. CRC Press, London.
16. Wadoux, A.M.J.C., & Brus, D.J. (2021). How to compare sampling designs for mapping? *European Journal of Soil Science*, *72*(1), 35–46. https://doi.org/10.1111/ejss.12962
17. We, X., & Ai, E. (2008). Top 10 algoritm in data mining. *Knowledge Information System*, *14*, 1-37.
18. Zhang, Y., Saurette, D.D., Easher, T.H., Ji, W., Adamchuk, V.I., & Biswas, A. (2022). Comparison of sampling designs for calibrating digital soil maps at multiple depths. *Pedosphere*, *32*(4), 588–601. https://doi.org/10.1016/S1002-0160(21)60055-3
19. Yang, L., Li, X., Shi, J., Shen, F., Qi, F., Gao, B., Chen, Z., Chen, A.X., & Zhou, C. (2020). Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method. *Geoderma*, *369*, 114337. https://doi.org/10.1016/j.geoderma.2020.114337
20. Yekom Consulting Engineers. (1988). Comprehensive studies on the revival and development of agriculture of the natural resources of the northern watershed of the Karun River.